The SARS Coronavirus: A Postgenomic Era

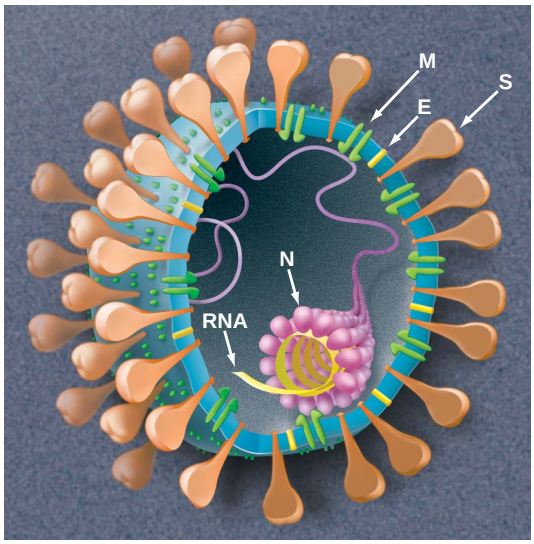
Kathryn V. Holmes and Luis Enjuanes

The complete sequences of the ~30,000-nucleotide RNA genomes of two isolates of the SARS coronavirus (SARS-CoV) are reported on pages 1399 and 1394 of this issue (*1*, *2*), a remarkable achievement since the virus was identified less than 2 months ago. Additional sequences in GenBank, and complete genome sequences of nine virus isolates from five patients (*3*) allow comparison between different SARS-CoV isolates.

Sequence analysis reveals the genome organization and phylogeny of SARS-CoV (*1*, *2*). The genome has all the features characteristic of a coronavirus, but is sufficiently different from all previously known coronaviruses to represent a new coronavirus group. The genomes of the SARS-CoV Tor2 strain from Toronto (*1*) and the Urbani strain from Vietnam (*2*) differ by just eight nucleotides. Thus, the viral RNA genome appears stable during human passage. The SARS-CoV genome contains five major open reading frames (ORFs) that encode the replicase polyprotein; the spike (S), envelope (E), and membrane (M) glycoproteins; and the nucleocapsid protein (N) in the same order and of approximately the same sizes as those of other coronaviruses. (The figure shows the virion's structure.)

Coronavirus genomes also contain a variable number of nonconserved ORFs interspersed between the major ORFs. Marra *et al.* identified nine potential ORFs, not found in other coronaviruses, that could encode proteins unique to SARS-CoV (*1*). Five of these were also identified by Rota *et al.* (*2*), who only included ORFs for proteins longer than 50 amino acids. It remains to be established which of these ORFs are translated in infected cells. These proteins may be nonessential for virus replication or they may serve novel functions in virus replication and pathogenesis or modulate immune responses to infection. The Marra and Rota groups proposed different names for these ORFs (*1, 2*) and a common nomenclature awaits experiments showing which ORFs are expressed in infected cells.

K. V. Holmes is in the Department of Microbiology, University of Colorado Health Sciences Center, Denver, CO 80262, USA. E-mail: kathryn.holmes@UCHSC.edu L. Enjuanes is in the Department of Molecular and Cell Biology, Centro Nacional de Biotecnologia, CSIC Campus Univ. Autonoma, Madrid, Spain. E-mail: L.Enjuanes@cnb.uam.es



CREDIT: KATHARINE SUTLIFF/*SCIENCE*

Coronavirus organization. A model of the coronavirus structure showing the organization of the spike (S), membrane (M), and envelope (E) glycoproteins. The RNA is protected by a helical capsid of N protein monomers.

Coronavirus-infected cells contain a characteristic 3′ coterminal nested set of mRNAs, each of which has at its 5′ end an ~70-nucleotide long, capped leader sequence derived from the 5′ end of the genome. Synthesis of subgenomic negative sense RNA species by discontinuous RNA transcription is regulated by a core transcription-regulating sequence (TRS) found on the genome near the beginning of each ORF and at the 3′ end of the leader (*4*). Marra *et al.* (*1*) suggest that the SARS-CoV TRS core sequence is 5′-CUAAAC-3′, like TRSs of murine and bovine coronaviruses in group 2. In SARS-CoV and human and porcine coronaviruses in group 1, this core sequence is flanked at its 3′ end by GAA. Rota *et al.* (*2*) suggest that the core TRS for SARS-CoV is 5′-AAACGAAC-3′, based on the 5′ sequence of the smallest mRNA. The TRSs of the known coronaviruses vary slightly from gene to gene, and candidate TRSs of SARS-CoV are not identical for each potential ORF. The consensus sequence CUAAAC is found before each of the S and M genes and ORF 10 of SARS-CoV, and the AAACGAAC sequence is found at the same three sites and also before the N gene and ORFs 3 and 9, suggesting that these six genes may be expressed in infected cells from subgenomic mRNAs. Sequences upstream of the E gene, and ORFs 7, 9, and 11 differ significantly from the consensus TRSs. Perhaps the E protein of SARS-CoV is translated from a larger mRNA by internal initiation, like the E proteins of several other coronaviruses. Rota and co-workers detected five abundant viral subgenomic mRNAs in Northern blots of cells infected with SARS-CoV, but less abundant mRNAs may not have been detected (*2*). Experimental data are needed to confirm the core TRS, characterize the viral mRNAs, and detect virus-encoded proteins in infected cells.

Although the predicted amino acid sequences of the 3CL protease, which is part of the viral replicase polyprotein, and the S, E, M, and N proteins of SARS-CoV suggest that they are structurally and functionally homologous to the proteins of known coronaviruses, the pairwise amino acid sequence identity with their homologs is less than 40 to 50%. Overall, the SARS-CoV genome appears to be equidistant from those of all known coronaviruses. The sequences of the polymerase are most closely related to bovine and murine coronaviruses in group 2, with some characteristics like avian viruses in group 3. In addition, the 3′ end of the SARS-CoV genome contains a 32-nucleotide motif that is also found in group 3 coronaviruses. Unlike group 2 viruses, SARS-CoV does not encode a hemagglutinin-esterase protein. Also, whereas group 2 viruses encode two papain-like proteinases in the replicase polyprotein, SARS-CoV, like group 3 viruses, apparently encodes a single papain-like proteinase. Based on comparison of the genomes of SARS-CoV and other coronaviruses, both the Rota and Marra groups suggest that SARS-CoV should be classified in a new coronavirus group.

Indeed, the genome clearly shows that SARS-CoV is neither a host-range mutant of a known coronavirus, nor a recombinant between known coronaviruses. SARS-CoV is also unlikely to have been created from known coronaviruses by genetic engineering, because at present it would be impossible to modify 50% of a coronavirus genome without abrogating viral infectivity. SARS-CoV probably evolved separately from an ancestor of the known coronavirus, and infected an unidentified animal, bird, or reptile host for a very long time before infecting humans and starting the SARS epidemic. The original host for the SARS-CoV may be identified by serological studies of species near the site

where the epidemic began. Coronavirus would have to be isolated from this host and its genome sequenced in order to identify genetic changes associated with adaptation to humans. The available sequence data on the few independent isolates of SARS-CoV from humans suggest that the virus is genetically quite stable. Minor nucleotide changes found in viruses from different clinical isolates may prove useful as markers for epidemiological studies, but their significance for viral pathogenesis cannot be determined until the functions and antigenic determinants of the viral proteins have been characterized.

The sequence of the SARS-CoV genome makes it possible to identify subgenomic mRNAs by reverse transcription–polymerase chain reaction and to clone viral cDNAs, express recombinant viral proteins, and study their roles in virus replication and pathogenesis. Viral cDNAs and antibodies to recombinant viral proteins will be useful for developing sensitive and specific tests for SARS-CoV RNA and antigens in clinical specimens. The genome sequence and recombinant viral proteins will also facilitate the development of drugs and vaccines against SARS-CoV. For example, a three-dimensional model of the SARS-CoV–encoded 3 CL proteinase, has been made to direct the design of protease inhibitors that may block coronavirus replication (*5*). Passive immunization with neutralizing monoclonal antibodies may be useful for prophylaxis or therapy. Live, attenuated vaccines may be developed by serial passage of SARS-CoV in cell culture, and mutations responsible for attenuation of virus virulence could then be identified. Characterization of the SARS-CoV antigens that elicit protective immunity will facilitate development of vaccines. Fortunately, coronavirus genomes can now be manipulated using targeted RNA recombination and infectious cDNA clones in order to identify determinants of virus virulence (*6*–*10*). Genetically engineered coronaviruses that can express proteins, but not be transmitted from cell to cell, may be useful as vaccines to elicit mucosal immunity (*11*–*13*). The direction of SARS research has now moved from identifying the virus and sequencing its genome to analyzing the viral proteins and their roles in virus replication and pathogenesis with the aim of developing new drugs and vaccines against SARS.

References

1. M. A. Marra *et al.*, *Science* **300**, 1399 (2003); published online 1 May 2003 (10.1126/science.1085953).
2. P. A. Rota *et al.*, *Science* **300**, 1394 (2003); published online 1 May 2003 (10.1126/science.1085952).
3. Y. J. Ruan *et al.*, *Lancet* **361**, published online 17 May 2003.
4. M. M. C. Lai, K. V. Holmes, in *Fields Virology*, D. Knipe *et al.*, Eds. (Lippincott-Williams and Wilkins, Philadelphia, 2001), vol. 1, pp. 1163–1185.
5. K. Anand *et al.*, *Science*, 13 May 2003 (10.1126/science.1085658).
6. P. S. Masters, *Adv. Virus. Res.* **53**, 245 (1999).
7. V. Thiel *et al.*, *J. Gen. Virol.* **82**, 1273 (2001).
8. R. Casais *et al.*, *J. Virol.* **75**, 12359 (2001).
9. F. Almazan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5516 (2000).
10. B. Yount *et al.*, *J. Virol.* **76**, 11065 (2002).
11. L. Enjuanes *et al.*, *J. Biotechnol.* **88**, 183 (2001).
12. J. Ortego *et al.*, *J. Virol.* **76**, 11518 (2002).
13. I. Sola *et al.*, *J. Virol.* **77**, 4357 (2003).

MATERIALS SCIENCE

Watching Nanoparticles Grow

Geoff Thornton

Because being very small can convey novel properties, there is much excitement about new methods of controlling and monitoring the growth of nanometer-scale particles. On page 1416 of this issue, Renaud *et al.* (*1*) report an important advance in this area. They show how the average shape and size of a supported nanoparticle ensemble can be followed during growth in real time (see the figure).

The unusual properties of nanoparticles are already employed in a number of applications. For example, TiO_2 particles in sunscreen maintain the beneficial ultraviolet-absorbing properties of the bulk material, but lose the cosmetically undesirable whitening as the particle size is decreased. Size and shape effects on the nanometer scale are also thought to play a crucial role in catalytic reactions involving supported metal particles (*2*). This is exemplified by the case of gold, which in its bulk form is inert, making it an ideal material for jewelry. By contrast, if the diameter of the particle is reduced to a few nanometers, it catalyzes the oxidation of CO to CO_2 (*3*). Many more applications can be envisioned. For instance, the size dependence of the band gap of semiconductors such as CdSe is being exploited in solar cell designs (*4*).

Both top-down and bottom-up approaches are being explored as a means of constructing nanometer-scale structures. The bottom-up approach of self-assembly is a particularly attractive synthetic route for nanoparticles because of its inherent simplicity and speed of production. For supported nanoparticles, competing methods include proximal probe (for example, scanning tunneling microscope, STM) manipulation (*5*) and deposition of clusters that have been preselected for size by a mass filter (*6*). As an example of top-down methodology, an ordered array of 40 nm-diameter Pt particles on silicon has been synthesized by first using electron beam lithography to write a pattern in a polymer resist (*7*).

Refinement of self-assembly methods requires tools for characterizing the size and shape of the nanoparticles. Scanning probe methods provide one means of achieving this goal. For metal nanoparticles on metal oxide substrates, one subject of the work of Renaud *et al.* (*1*), individual surface atoms can be resolved with a STM (*8*). Furthermore, STM allows the electronic structure to be investigated in parallel with the topography with scanning tunneling spectroscopy (*9*). However, STM does have weaknesses. Tip shadowing effects make it difficult to follow the growth of nanoparticles in situ. Moreover, the technique is usually only sensitive to the top layer, so information about the core of the nanoparticle is absent.

Three-dimensional (3D) analysis of nanometer-scale structures is possible with transmission electron microscopy (TEM) tomography. In some ways this is akin to the operation of a medical computed tomography scanner, with TEM images collected at different angles of sample tilt in the microscope stage. The resulting computer-generated 3D images typically have nanometer resolution, leading to the widespread use of the technique in the analysis of macromolecules. It is more difficult to work with crystalline materials. Nevertheless, impressive results have been achieved on catalyst particles (*10*).

TEM has also been used to examine the morphology of nanoparticles as a function of background gases in the millibar range and at elevated temperatures. Cu nanocrystals on a ZnO support imaged at atomic resolution were observed to change shape with gas composition (*11*). In 1.5 mbar of H_2, the angular shape of the nanocrystals is similar to that observed under ultrahigh-vacuum conditions. Addition of water vapor causes the particle to adopt a more spherical shape, because water is adsorbed on the exposed facets. This shape change is reversed when the gas mixture is replaced by pure H_2.

Coherent x-ray diffraction methods have also been used to image nanostructures (*12*). In principle, diffraction methods have the advantage of being able to sample relatively large areas at high spatial resolution. Coherent x-ray beams are generated by third-generation synchrotron radiation sources. For instance, the Advanced Photon Source at Argonne National Laboratory has

The author is in the Nanoscience Research Centre and Chemistry Department, Manchester University, Manchester M13 9PL, UK. E-mail: g.thornton@man.ac.uk